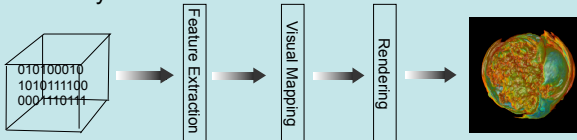


An Information-Theoretic Framework for Enabling Extreme-Scale Science Discovery

Project Overview

- Overall research goals
 - Create a quantitative data analysis model to analyze the information flow across the entire data analysis and visualization pipeline
 - Minimize the information losses and maximize the quality of analysis
- Specific aims
 - Decide what data are the most essential for analysis
 - Transform data into effective representations/visualizations that rapidly convey the most insight
- An Information-theoretic approach
 - Quantify Information content using various entropy measures
 - Steer the analysis of data based on information saliency



What is Information Theory?

- Study the fundamental limits to reliably transmitting messages through a noisy channel
- Model the message as a random variable whose value is taken from a sequence of symbols
- Information content can be measured by Shannon's Entropy

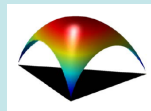
Shannon's Entropy

The random variable takes a sequence of symbols $\{a_1, a_2, a_3, \dots, a_n\}$ with probabilities $\{p_1, p_2, p_3, \dots, p_n\}$. The information content of each symbol a_i is defined as

$$\log(1/p_i) = -\log p_i$$

The average amount of information expressed by the random variable is

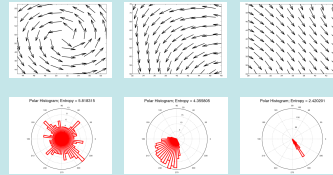
$$H(x) = -\sum_{i=1}^n p_i \log p_i$$



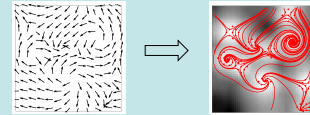
An entropy function of 3 symbols

Visual Analytics Example

Step1: local entropy measurement - $H(X)$



Step2: Initial streamline seeding - $H(Y)$



Step3: Refinement of streamline seeding - minimize $H(X|Y)$

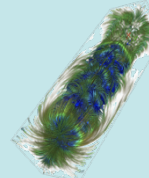
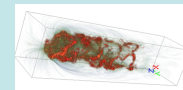
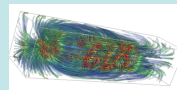
Vector Field X



Streamlines Y



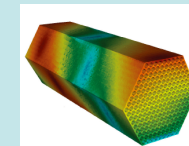
Entropy-based visual mapping



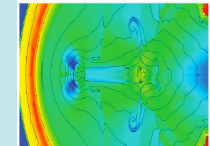
Science Applications

- Nek5000: A Navier-Stokes solver for fluid flow, convective heat, and magnetohydrodynamics simulations
- Flash: Adaptive mesh code for astrophysics and cosmology

Our collaborators demand large-scale, high-fidelity visual analysis. The snapshots below show examples of their current visualizations of single quantities, but scientists are interested in the simultaneous analysis and visualization of all of the quantities being computed, both in situ and during post-processing.



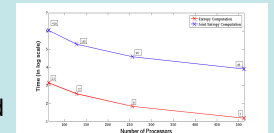
Coolant flow around fuel pins in a nuclear reactor



Temperature in a collision of galaxy cluster

ITL Software

- Information-Theoretic Library (ITL)
- Entropy analysis for exascale data sets
- Seamlessly integrated into large-scale simulations to provide in situ data analysis and visualization.



Performance of ITL run on NERSC's Franklin (Cray XT4). Our initial results showed that satisfactory scalability can be achieved.

Next Steps

- Information Analysis for Multi-resolution Data
 - Structured and unstructured data LODs
- Time-varying Multivariate Analysis
 - Temporal trend analysis and indexing
- Information-Aware Visual Analysis